

Editorial

On the Correct Use of 'Data'

Dmitry N. Tychinin^{1*} and Almut Beate Heinrich²

¹ Institute of Biochemistry and Physiology of Plants and Microorganisms, Russian Academy of Sciences, 13 Prospekt Entuziastov, Saratov 410049, Russian Federation (tychinin@ibppm.sgu.ru)

² Publisher editor, Scientific Journals, ecomed publishers, Justus-von-Liebig-Str. 1, D-86899 Landsberg, Germany (a.heinrich@ecommed.de)

* Corresponding author

ED BEESE'S [1] article upholding the use of 'data' as a singular noun elicited the following comment from Dr. PAUL JOHNSTON (Greenpeace International Research Laboratories, Dept. of Biological Sciences, University of Exeter):

In scientific, philosophical and general use, this word is usually considered to denote a number of items and is thus treated as a plural with 'datum' as the singular. In computing and allied subjects (and sometimes in general use) it is treated as a mass (or collective) noun and used with words like 'this', 'that', and 'much', with singular verbs, e.g. 'useful data has been collected' (*The Concise Oxford Dictionary*).

Dr. JOHNSTON pointed out that 'datum', the singular form of 'data', is still in common use, especially as part of 'datum point', meaning a fixed reference point in, *inter alia*, archaeology, construction, or surveying.

Because the correct use of 'data' is very important in the Int J LCA context, we thought it was worthwhile to take a look at how the word is treated by authorities on English style and usage. Here is what we found.

Before the advent of computers, 'data', an English word of Latin origin, was considered a plural noun having 'datum' as its singular form. For example, *Chambers's Twentieth Century Dictionary of the English Language*, 1925 [2], defines 'data' thus:

Data, *n. pl.* facts given or admitted from which other facts may be deduced: – *sing.* Datum. [*L. datum, data, given – dāre, to give.*]

After World War Two, the singular use of 'data' became widespread in computing and also in ordinary speech – showing possibly the great excitement among people over advances in computer technology. The word's meaning expanded to include 'information' (uncountable) as well as 'facts' (countable). Since then, language experts on both sides of the Atlantic have been debating the legitimacy of the new usage. "The use of 'data' as if it were a singular noun is a common solecism," claims the former *New York Times* editor THEODORE BERNSTEIN in his *The Careful Writer*, 1965 (quoted by Day [3]); but the British novelist KINGSLEY AMIS [4] remarks, "Yes, by origin plural only, by usage irreversibly singular." The British *Longman Guide to English Usage* [5], however, notes that in formal writing, it is better to treat the word as plural; and *Webster's Ninth New Collegiate Dictionary* (American) says plural use is much more common than singular use.

Of particular interest to us were comments from two of the world's leading experts on scientific writing. According to MICHAEL ALLEY [6]:

The word *data* is a plural form of *datum*, a Latin word. Because *datum* is seldom used as the singular form, many sources consider *data* acceptable as either singular or plural. Some conservatives, though, refuse to accept this word as singular. If you need a singular form and do not want to distract your conservative audience, spend an extra word and write *data point*.

ROBERT DAY [3], a professor of English and a former managing editor of the *Journal of Bacteriology* (American Society for Microbiology), observes:

The problem is severe in scientific writing, especially in biology, because so many of our words are, or derived from, Latin. Most of these words retain their Latin plurals; at least they do so when used by careful writers. Many of these words (e.g., *data*, *media*) have entered popular speech, where the Latin "a" plural ending is simply not recognized as a plural. Most people habitually use "data is" constructions and probably have never used the real singular, *datum*. Unfortunately, this lax usage has become so common outside science that even some dictionaries tolerate it.

So, the confusion is not so much between Latin and English as it is between scientific and unscientific use of the word. Although the modern tendency is to treat 'data' as a singular noun ('data is available'), in scientific use – particularly in biology, which owes much more to Latin than does computing or everyday language – we should keep the tradition alive and treat 'data' (= 'facts', 'values') as a plural noun ('data are available'). Let us also remember that the key characteristic of science writing is clarity [3] and that 'data is' constructions are likely to confuse many foreign speakers of English. Until singular use supersedes plural use in scientific discourse, 'data are' constructions will remain familiar and unambiguous.

The discussion is not closed, and your opinions are very welcome: Shall we at Int J LCA adopt a 'data are' policy?

Acknowledgement. Thanks are due to Drs. Paul Johnston (P.Johnston@exeter.ac.uk) and E. William Beese (ebeese@t-online.de) for kindly providing dictionary definitions and for sharing their thoughts about the correct use of 'data'.

References

- [1] Beese EW (2003): English language usage: Frequent errors by non-native speakers and how to avoid them. *Environ Sci & Pollut Res* 10: 205
- [2] Davidson T (Ed) (1925): *Chambers's Twentieth Century Dictionary of the English Language*. W. & R. Chambers, Ltd., London and Edinburgh
- [3] Day RA (1998): *How to Write and Publish a Scientific Paper*, 5th ed. Oryx Press, Phoenix, Ariz
- [4] Amis K (1998): *The King's English. A Guide to Modern Usage*. HarperCollins, London
- [5] Greenbaum S, Whitcut J (1990): *Longman Guide to English Usage*, Spec. ed. Russky Yazyk Publishers, Moscow
- [6] Alley M (2000): *The Craft of Editing: A Guide for Managers, Scientists, and Engineers*. Springer-Verlag, New York